Microsoft

The Cortana Intelligence Suite Foundations

Process and Platform, Environment Configuration

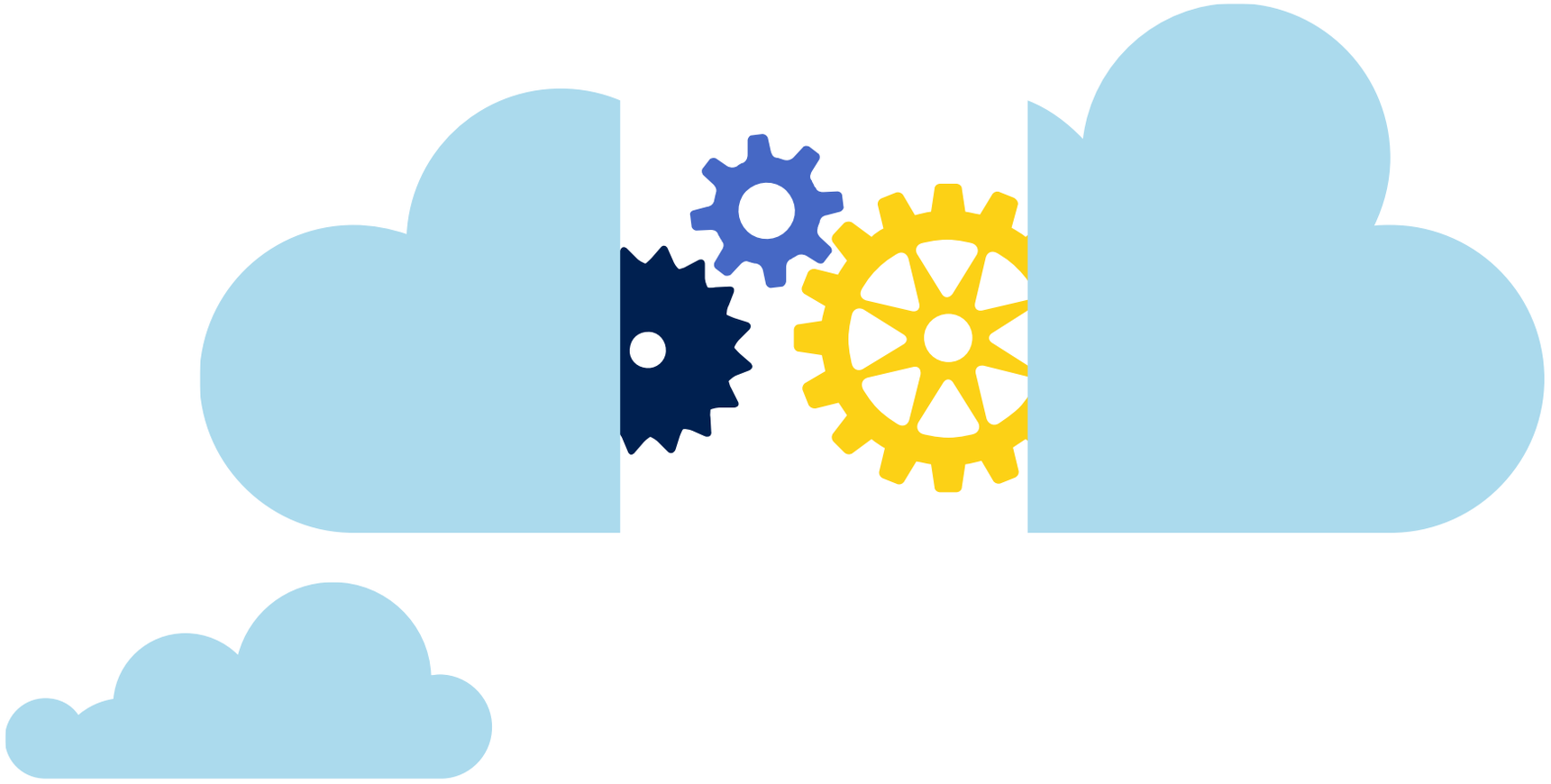
Mithun Prasad, PhD
miprasad@Microsoft.com

Session	Concepts	Technologies
Process and Platform Environment Configuration	The CIS Process, CIS Platform components, Tools installation and overview	CRISP-DM, CIS, Azure Portal, ADC Interface, Visual Studio Interface (and RTVS), Power BI Interface, Azure Machine Learning Interface, Azure PowerShell, Azure Storage Explorer
Data Discovery and Ingestion	Data sourcing, Feature selection techniques, Data cataloging, Data Ingestion, Data Exploration	Azure Data Catalog, Azure Storage, Techniques for discovery
Data Preparation	Data selection, including Features, Dimension reduction, Data processing, Data transformation and augmentation	Azure Data Factory, HDInsight
Modeling for Machine Learning and Data Mining	Algorithm selection and application, Parameter selection and adjustment	Azure Machine Learning, Microsoft R Server overview, Azure Data Factory
Business Validation and Model Evaluation	Business validation of report and results, Model testing and cross-validation	Azure Machine Learning, Microsoft R Server overview, Azure Data Factory, Business Validation, SQL DB, Azure Storage
Deploying and Accessing the Solution	Deploying the solution using Data Destinations, Deploying the solution using API's, Deploying the Solution using Queries and Reports	Azure Data Storage, SQL DB, Azure Machine Learning API, Cognitive Services API, HIVE, Power BI
Workshop recap	Mapping requirements to solution elements, What to use When	Understand when to use each component within CIS

Learning Objectives

- Understand the CIS Process
- Understand the CIS Platform
- Set up your Development Environments

Module 1: Understanding Cortana Intelligence



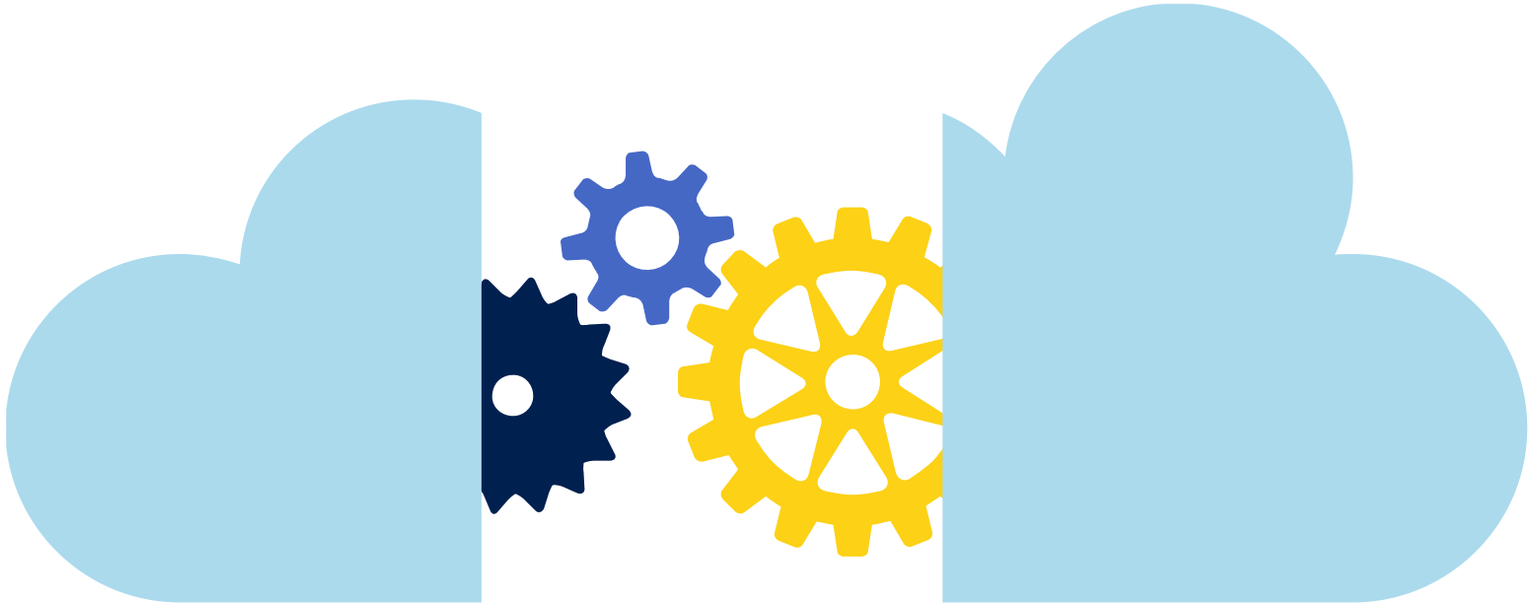
Microsoft

Cortana Intelligence in a Sentence

Cortana Intelligence is a Platform and a Process to perform advanced analytics from start to finish

Microsoft

Data Science Process and Platform



Microsoft

The Team Data Science Process

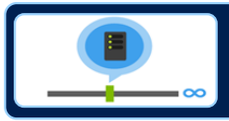
Business Understanding	• Define Objectives • Identify Data Sources
Data Acquisition and Understanding	• Ingest Data • Explore Data • Update Data
Modeling	• Feature Selection • Create and Train Model
Deployment	• Operationalize
Customer Acceptance	• Testing and Validation • Handoff • Re-train and re-score

The Cortana Intelligence Platform

Cortana, Cognitive Services, Bot Framework

Power BI

Stream Analytics



HDInsight

Azure Machine Learning (MRS)

SQL Data Warehouse (SQL DB, Document DB)

Data Lake

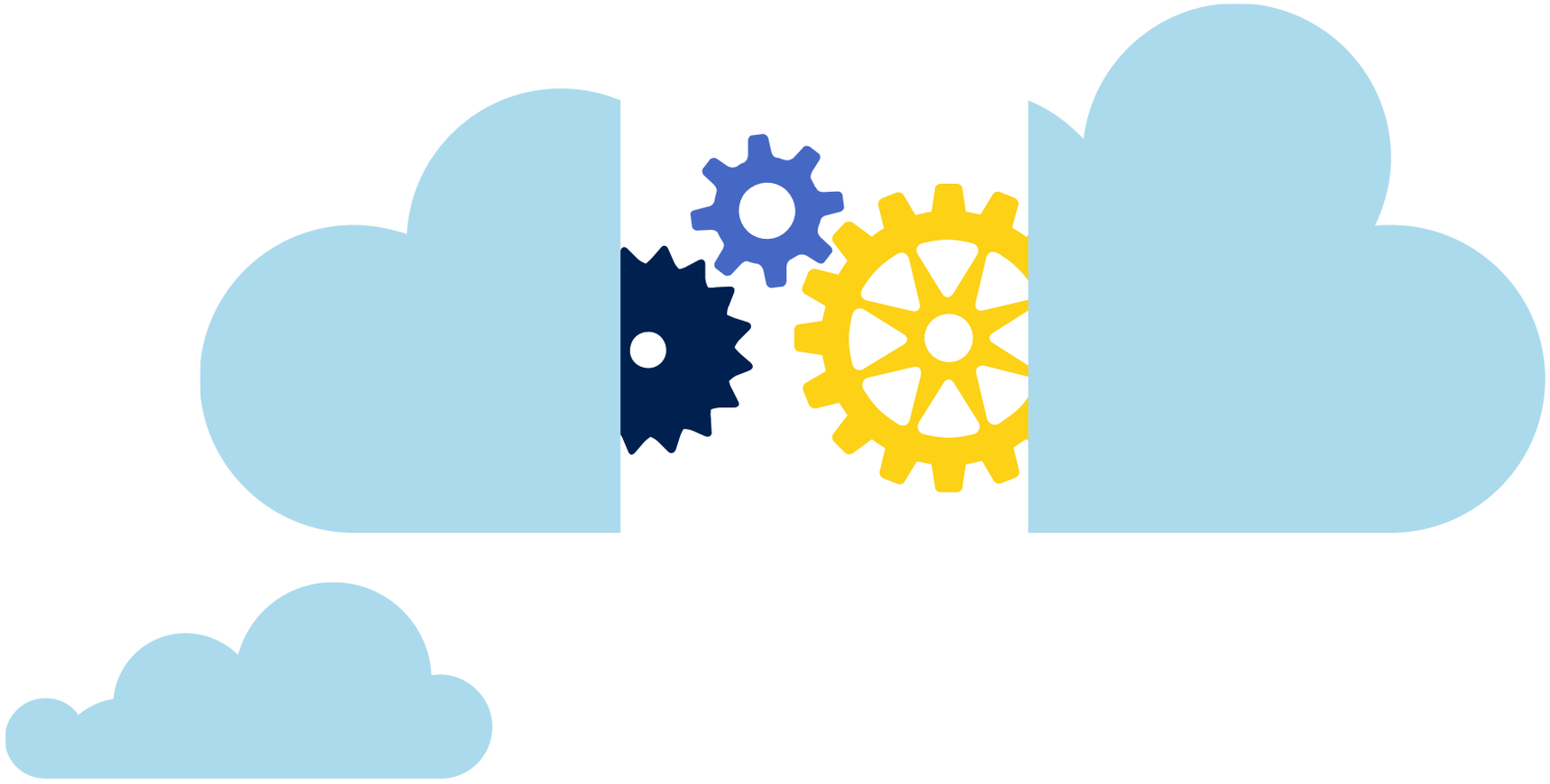
Event Hubs

Data Factory

Data Catalog

Microsoft Azure

Module 2: Cortana Intelligence Suite



Microsoft

Microsoft Azure

What it is:

Microsoft's Cloud Platform including IaaS, PaaS and SaaS

When to use it:

- Storage and Data
- Networking
- Security
- Services
- Virtual Machines
- On-demand Resources and Services

Microsoft

Azure Data Catalog

What it is:

On-Line Catalog of Meta-Data about your Data Sources, with easy tagging and searching

When to use it:

- Sourcing data
- Data discovery
- Data vetting

Microsoft

Azure Data Factory

What it is:

A pipeline system to move data in, perform activities on data, move data around, and move data out

When to use it:

- Create solutions using multiple tools as a single process
- Orchestrate processes - Scheduling
- Monitor and manage pipelines
- Call and re-train Azure ML models

Microsoft

Event Hubs

What it is:

A system to ingest data from the web, IoT, and apps at scale

When to use it:

- To stream in large amounts of data
- With IoT workloads
- Use with variable or unpredictable large data loads
- Similar to Kafka

Microsoft

Data Lake

What it is:

Data storage (Web-HDFS) and Distributed Data Processing (HIVE, Spark, HBase, Storm, U-SQL) Engines

When to use it:

- Low-cost, high-throughput data store
- Non-relational data
- Larger storage limits than Blobs

DocumentDB

What it is:

An automatically-indexed, schema-agnostic JSON database

When to use it:

- Query non-relational data
- Schema defined per object
- Document (JSON) – Oriented database
- Ad-hoc queries
- Stored Procedures

Microsoft

SQL Database

What it is:

A SQL Server Database Service in the Cloud

When to use it:

- When you need a relational store
- When you need full transactional support
- When you have familiarity with SQL and T-SQL and SQL Server Objects
- When you need lots of flexible indexing
- When you do not want to manage a SQL Server
- When you have multitenant databases needed

Microsoft

SQL Data Warehouse

What it is:

A Scaling Data Warehouse Service in the Cloud

When to use it:

- When you need a large-data BI solution in the cloud
- When you are using lots of relational data
- When you need lower cost relational storage than Blobs
- When you need pause-able scaled compute

Microsoft

Azure ML

What it is:

A multi-platform environment and engine to create and deploy Machine Learning models and API's

When to use it:

- When you need to create predictive analytics
- When you need to share Data Science experiments across teams
- When you need to create call-able API's for ML functions
- When you also have R and Python experience on your Data Science team

Microsoft

Microsoft R Server (MRS)

What it is:

A scalable, highly-performing R engine used in on-prem, in-cloud, and in-service areas

When to use it:

- When you need to use the R language and environment for data processing at scale

Microsoft

HDInsight

What it is:

Microsoft's implementation of apache Hadoop (as a service) that uses Blobs for persistent storage

When to use it:

- When you need to process large scale data (PB+)
- When you want to use Hadoop or Spark as a service
- When you want to compute data and retire the servers, but retain the results
- When your team is familiar with the Hadoop Zoo

Microsoft

Stream Analytics

What it is:

Real-time cloud-based stream processing

When to use it:

- For complex event processing
- IoT, streaming workloads
- When you to ingest need millions of records per second
- When you need JSON, Delimited, and Avro data processing
- Similar to Apache Storm

Microsoft

Power BI

What it is:

Interactive Report and Visualization creation for computing and mobile platforms

When to use it:

- When you need to create and view interactive reports that combine multiple datasets
- When you need to embed reporting into an application
- When you need customizable visualizations
- When you need to create shared datasets, reports, and dashboards that you publish to your team

Microsoft

Cortana and Cognitive Services, Bot Framework

What it is:

Intelligent assistant available in computing and mobile platforms, integrated into user's ecostructure, speech and vision interaction

When to use it:

- When you want your users to interact with your solution in a natural language format
- When you have an application of your solution that lends itself to the user's connected ecostructure

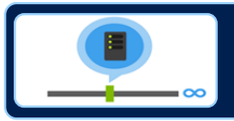
Microsoft

The Cortana Intelligence Platform

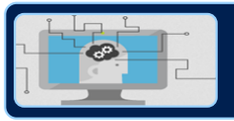
Cortana, Cognitive Services, Bot Framework

Power BI

Stream Analytics



HDInsight



Azure Machine Learning (MRS)

SQL Data Warehouse (SQL DB, Document DB)

Data Lake

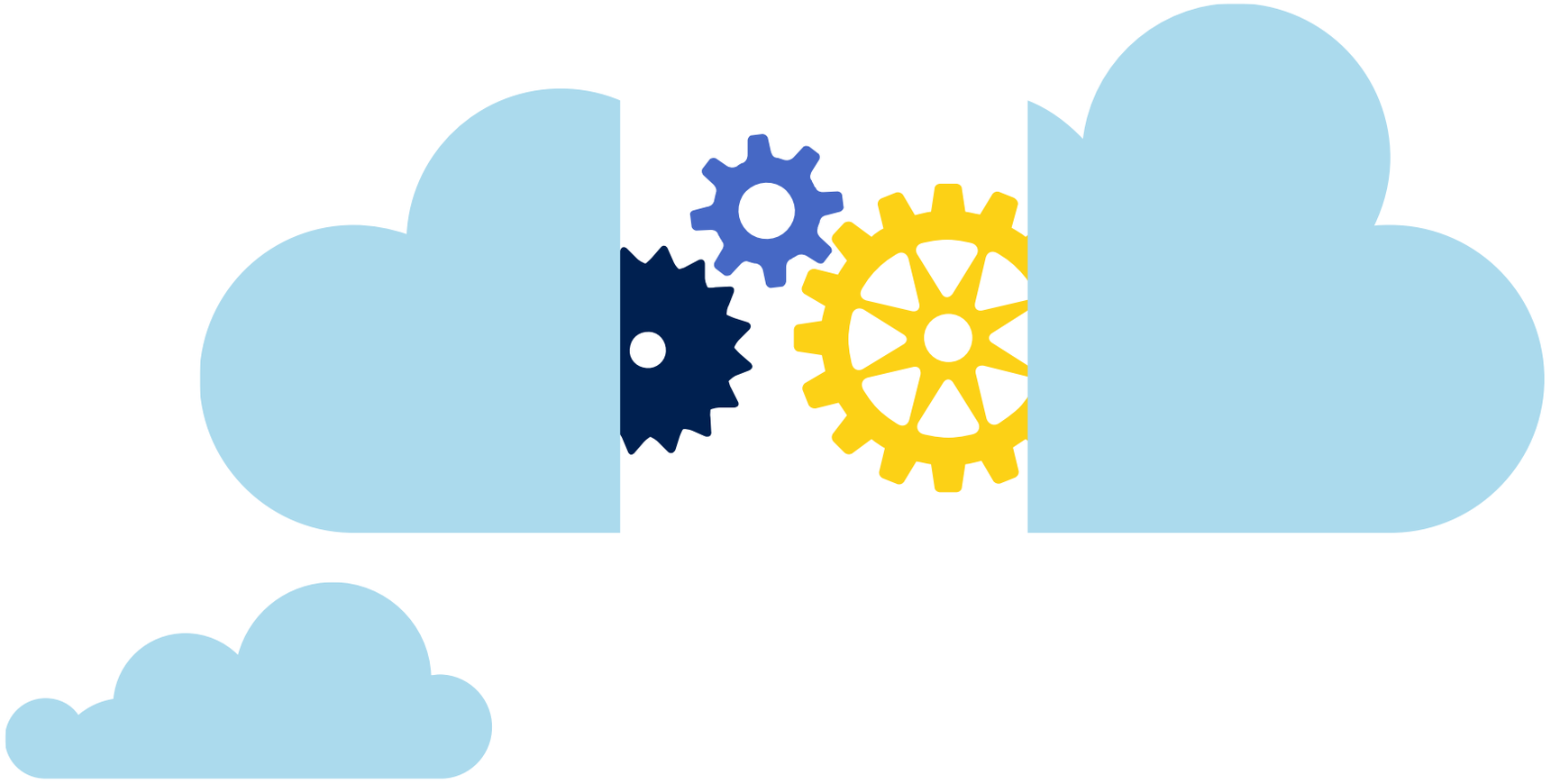
Event Hubs

Data Factory

Data Catalog

Microsoft Azure

Module 3:
Setting up Your Development Environment



Microsoft

Primary Development Tools:

- The Azure Portal
- Azure SDK
- Azure PowerShell and ARM Templates
- Azure Data Catalog
- Azure ML Interface
- Visual Studio (and RTVS)
- Storage Explorer

Microsoft

Skills check

You should now be able to:

- Understand the CIS Process
- Understand the CIS Platform
- Set up and Configure your Development Environments